المقارنة بين دالتي بقاء تتوزعان اُسياً

زينب يوسف داود *

المستخلص :

هدفت هذه الدراسة الى المقارنة بين زمن بقاء المرضى المصابين بسرطان الرئة والمرضى المصابين بسرطان المعدة حيث تم سحب عينة عشوائية بحجم 30 مشاهدة من سجلات المرضى الذين رقدوا في دار التمريض الخاص خلال الفترة 1-6-1999 ولغاية 31-12-1999 اذ وجد إن بيانات كل مجموعة من المرضى تتبع التوزيع الاسي باستخدام اختبارNelson بطريقة رسم معدل الخطورة التجميعية وعلى هذا الاساس تم استخدام اختبار نسبة الامكان الاعظم واختبار cox F المعلمية للبيانات المراقبة واثبت كلا الاختباريين عدم وجود فروق معنوية لكلا المجموعتين .

المقدمة :

تعد دراسات البقاء من الدراسات الإحصائية المهمة التي تهتم بدراسة التجارب الطبية فكثيرا من الأحيان قد يجد الباحث نفسه أمام مقارنة بين مجموعتين من المرضى فاقترح الباحثون بهذا المجال العديد من الاختبارات المعلمية و اللامعلمية, فعندما تتوفر لدى الباحث بيانات لمجموعتين مستقلتين من المرضى كل منها تتبع توزيع معلمي معروف يلجا إلى اختبارها وفق اختبار نسبة الامكان الأعظم Likelihood ratio test سواء كانت البيانات مراقبة أو غير مراقبة ولكون هذا الاختبار لا يخلو من نقاط ضعف إذ يكون مقربا إلى توزيع كاي سكوير و يتطلب حجم عينة لا يقل عن 25 مشاهدة لذا اقترح الباحث Cox عام 1953 اختباراً للمقارنة بين دالتي بقاء تتوزعان اسياً عندما تكون البيانات مراقبة أكثر قوة من السابق إذ يكون مطابقا إلى توزيع F .

* مدرس مساعد / الجامعة المستنصرية / كلية الادارة والاقتصاد / قسم الاحصاء

مقبول النشر بتاريخ 2005/3/23

هدف البحث

هناك هدفان أساسيان لهذا البحث هما

1.المقارنة بين مجموعتين مستقلتين غير متماثلتين بيانات كل منها مراقبة و تتبع التوزيع الاسي.

2.الكشف عن مطابقة البيانات للتوزيع الاسي عندما تكون البيانات مراقبة بطريقة الرسم.

الجانب النظري

1- اختبار نسبة الامكان الأعظم Likelihood ratio test

هو أسلوب لاختبار الفرضيات مقابل بديلتها باستخدام مقدرات الامكان الأعظم لعينتين مستقلتين لهما توزعي بقاء S1(t) و S2(t)

ولا يشترط أن يكون حجمي العينتين متساويين ويمكن صياغة الاختبار كالأتي

لتكن $x_1.x_2\ldots x_n$ عينة عشوائية بحجم n من مجتمع له دالة توزيع تكراري $f(x;\theta_1,\theta_2,\ldots,\theta_k),n>k$ والأساس لهذه العينة يتطلب اختبار الفرضية r ل k من المعالم $\theta_1,\theta_2,\ldots\theta_k$

إنها متطابقة $H_0:r$

جميع هذه المعالم غير متطابقة : H_a

ليكن w يمثل أبعاد فضاء عينة جزئية (k-r)من المعالم الغير متطابقة تحت فرضية العدم H_0.

ليكن Ω يمثل أبعاد فضاء عينة جزئية k من المعالم الغير متطابقة تحت فرضية H_a

افترض أن $\theta_1,\theta_2,\ldots\theta_r$ معالم متطابقة تحت فرضية العدم H_0 وتأخذ القيم $\theta_1{}'',\theta_2{}'',\ldots\theta_r{}''$

تستبدل باقي المعالم الغير متطابقة $\theta_{r+1},\theta_{r+2},\ldots\theta_k$بمقدرات الإمكان الأعظم $\hat{\theta}_{r+1},\hat{\theta}_{r+2},\ldots\hat{\theta}_k$فان دالة الإمكان الأعظم $L(\hat{w})$ تكون كالأتي

$$L(\hat{w})=\prod_{i=1}^{n}f(x_i;\theta_1{}'',\ldots,\theta_r{}'';\hat{\theta}_{r+1},\ldots,\hat{\theta}_k)$$

وتحت الفرضية H_a جميع المعالم تستبدل بمقدرات الإمكان الأعظم لذلك فان دالة الإمكان الأعظم $L(\hat{\Omega})$ تكون كالآتي

$$L(\hat{\Omega}) = \prod_{i=1}^{n} f(x_i;\hat{\theta}_1,\dots,\hat{\theta}_k)$$

و احصاءة الاختبار Λ تكون كالآتي

$$\Lambda = \frac{L(\hat{w})}{L(\hat{\Omega})} \qquad \dots 1$$

وليس من الصعب معرفة $0 < \Lambda < 1$ إذ أن $L(\hat{w})$ و $L(\hat{\Omega})$ هما دوال إمكان أعظم وان النسبة تكون موجبة ولان $L(\hat{w})$ مقيدة بحد أعلى وان $L(\hat{\Omega})$ غير مقيدة بحد أعلى لذلك فان $L(\hat{w}) < L(\hat{\Omega})$ وهذا يجعل $\Lambda < 1$. والقيمة $-2Log_e\Lambda$ تقترب من توزيع χ^2 بدرجة حرية r .

2.تقدير معلمة التوزيع الاسي للبيانات المراقبة

لدالة الوفاة التي لها التوزيع التكراري الآتي

$$f(t) = \lambda\exp(-\lambda t) \qquad \lambda > 0, t \ge 0$$

احتمال أن تموت المفردة في فترة الدراسة يكون كالآتي

$$1 - S(t) = \int_0^{T_i} \lambda\exp(-\lambda\, t)dt = 1 - \exp(-\lambda\, Ti)$$

إذ أن T_i يمثل زمن الدراسة الكلي

ليكن $\mu = \lambda^{-1}$

نلاحظ إن

$$f(t_i) = \begin{cases} \mu^{-1}\exp(-\mu^{-1}t_i) & 0 \le t_i \le T_i \\ \exp(-\mu^{-1}T_i) & t_i > T_i \end{cases} \qquad \dots\dots 2$$

وهذا لان المفردة إما تموت خلال الفترة $t_i \le T_i$ بدالة كثافة احتمالية $\mu^{-1}\exp(-\mu^{-1}t_i)$ أو إنها تبقى على قيد الحياة ما بعد الفترة T_i وبهذا نستطيع فقط قياس احتمال بقاءها أي $\exp(-\mu^{-1}T_i)$ أو $S(T_i)$ وبهذا فان دالة الإمكان الأعظم تكون كالآتي :

$$L(\mu) = \prod_{i=1}^{n} \left[f(t_i)\right]^{\delta_i} \left[S(T_i)\right]^{1-\delta_i} \quad ......3$$

إذا توفي المريض خلال الفترة

إذا لم يتوفى خلال الفترة

$$\boxed{\delta = \begin{cases} 1 & 0 \le t_i \le T_i \\ 0 & 0 \le t_i \le T \end{cases}}$$

وعند تعويض المعادلة رقم 2بالمعادلة رقم 3 واخذ اللوغاريتم لها نحصل على المعادلة آلاتية :

$$Log_e L = -\sum_{i=1}^{n} \left[\delta_i \left(Log_e \mu + \mu^{-1} t_i\right) + \left(1-\delta_i\right)\mu^{-1} T_i\right]$$

وعند وضع $\partial Log_e L / \partial \mu /_{\mu=\hat{\mu}} = 0$ نحصل على

$$\sum_{i=1}^{n} \left[\delta_i(-\hat{\mu}^{-1} + \hat{\mu}^{-2} t_i) + (1-\delta_i)\hat{\mu}^{-2} T_i\right] = 0$$

ومنها نحصل على

$$\hat{\mu} = d^{-1} \sum_{i=1}^{n} \left(\delta_i t_i + (1-\delta_i) T_i\right) \qquad ....4$$

حيث إن $d = \sum_{i=1}^{n} \delta_i$ والتي يفترض أن تكون اكبر من الصفر و $\hat{\mu}$ المعطاة في المعادلة 4 هي مجموع أوقات بقاء المرضى حتى حدوث الوفاة أثناء الدراسة مضافاً إلى مجموع أوقات بقاء المرضى الذين بقوا على قيد الحياة إلى ما بعد فترة الدراسة ومنها نحصل مقدر الإمكان الأعظم λ كالأتي

$$\hat{\lambda} = \frac{d}{\sum_{i=1}^{n} (\delta_i t_i + (1-\delta_i) T_i)} \qquad ......5$$

إذ أن d هي عدد المتوفيين فإذا كان d=0 فان $\hat{\mu} = \sum_{i=1}^{n} T_i$

3ـ اختبار نسبة الإمكان الأعظم للتوزيع الاسي لبيانات مراقبة

لتكن n_1 و n_2 عينتين عشوائيتين من المرضى مستقلتين غير متماثلتين لكل منها دالة بقاء

$$S_j(t,\lambda_j)=\exp(-\lambda_j t) \quad j=1,2$$

دالة الإمكان الأعظم

$$L(\Omega)=\lambda_1^{d_1}\lambda_2^{d_2}\exp\left(-\sum_{j=1}^{2}\sum_{i=1}^{n_j}\lambda_j\left\{\delta_{ij}t_{ij}+(1-\delta_{ij})T_{ij}\right\}\right) i=1,\ldots n; j=1,2 \qquad \ldots 6$$

وتحت فرضية العدم $H_0:\lambda_1=\lambda_2$ فان

$$L(w)=\lambda^{d_1+d_2}\exp\left(-\lambda\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left\{\delta_{ij}t_{ij}+(1-\delta_{ij})T_{ij}\right\}\right) \quad \ldots\ldots 7$$

مقدرات الإمكان الأعظم $\hat{\lambda}_j$. $\hat{\lambda}$ على التوالي كما ذكر في المبحث السابق

$$\hat{\lambda}_j=\frac{d_j}{\sum_{i=1}^{n_j}\left(\delta_{ij}t_{ij}+(1-\delta ij)T_{ij}\right)} \qquad \ldots\ldots 8$$

$$\hat{\lambda}=\frac{d_1+d_2}{\sum_{j=1}^{2}\sum_{i=1}^{m}\left(\delta_{ij}t_{ij}+(1-\delta ij)T_{ij}\right)} \quad j=1,2 \qquad \ldots\ldots 9$$

ولاختبار الفرضية

$$H_0:\lambda_1=\lambda_2$$
$$H_a:\lambda_1\neq\lambda_2$$

نقوم باحتساب الاحصاءة Λ كما في المعادلة 1 وكما ذكر سابقاً $-2Log_e\Lambda$ تقترب من توزيع كاي سكوير بدرجة حرية 1 إذ ترفض فرضية العدم إذا تجاوزت قيمة $-2Log_e\Lambda$ قيمة كاي سكوير الجد ولية بمستوى معنوية α.

4.اختبار Cox F

اقترح Cox عام 1953 اختباراً للفرضية $H_0 : \lambda_1 = \lambda_2$ مقابل بديلتها من جانب واحد أو جانبين لعينتين عشوائيتين مستقلتين تتبعان التوزيع الاسي كالآتي

ليكن

$$t_{ij} ; i = 1,2,\ldots,d_j \qquad \text{و} \qquad T_{ij} ; i = d_{j+1},\ldots,n_j \qquad j = 1,2$$

أوقات بقاء كما تم تعرفها مسبقاً فإذا كانت البيانات مراقبة فان النسبة $\frac{\bar{t}_1}{\bar{t}_2}$ تطابق توزيع F بدرجة حرية $2d_1$ و $2d_2$ للبسط والمقام على التوالي إذ أن

$$\bar{t}_j = \hat{\lambda}_j^{-1} \quad j = 1,2$$

فان تجاوزت قيمة F المحتسبة مع قيمة F الجد ولية بمستوى معنوية α عندئذ لا يمكن قبول فرضية العدم .

5. اختبار Nelson

عندما تتبع بيانات الدراسة التوزيع الاسي فان معدل الخطورة λ يكون ثابتاً وبهذا يمكن الحكم على أن البيانات قيد الدراسة تتبع التوزيع الاسي من خلال رسم معدل الخطورة ، ولكن في بعض الأحيان تأثير بعض المفردات تجعل معدل الخطورة غير منتظم وبهذا لا يمكن تمييز معدل الخطورة ، وللتغلب على هذه المشكلة اقترح Nelson عام 1972 رسم معدل الخطورة التجميعي إذ أن معدل الخطورة التجميعي للتوزيع الاسي يكون كالآتي

$$\Lambda(t) = \int_0^t \lambda \, dx = \lambda \, t \qquad t \geq 0$$

ويمكن توضيح الأسلوب الذي اقترحه Nelson بالخطوات آلاتية

1- تُرتب المشاهدات (n) تصاعدياً من أصغر قيمة إلى أكبر قيمة سواء كانت مراقبة أو غير مراقبة مع وضع علامة للمشاهدة المراقبة.

2- تحسب قيمة المخاطرة لكل الأوقات كالأتي $100/k$ إذ أن k تمثل مقلوب الرتبة.

3- نقوم بترقيم المحور العمودي من الصفر إلى اكبر قيمة لأوقات البقاء والمحور الأفقي من الصفر إلى اكبر قيمة من قيم الخطورة التجميعية.

4- ترسم أوقات البقاء على المحور العمودي مقابل قيم الخطورة التجميعية.

وبالنظر يمكن الحكم على مطابقة البيانات للتوزيع فإذا كانت النقاط متجه و منعقدة حول خط المستقيم المار بالصفر وغير متباعدة عنه يمكن الحكم على إن البيانات تتبع التوزيع الاسي.

الجانب العملي

يهدف المقارنة بين المرضى المصابين بسرطان المعدة و المرضى المصابين بسرطان الرئة لخطورتهما العالية وفق ما تشير إليه الملاحظات الأولية تم سحب عينة عشوائية بحجم 30 مفردة من سجلات المرضى الراقدين في مستشفى دار التـمريض الخاص لتفادي الضعف الممكن أن يحصل في اختبار نسبة الإمكان الأعظم للفترة من 1-6-1999 ولغاية 31-12-1999 وقمنا بإجراء الاختبارات الإحصائية المذكورة بالجانب النظري.

نقوم أولا بمطابقـة البيانات للتوزيع الاسي وفق أسلوب Nelson .

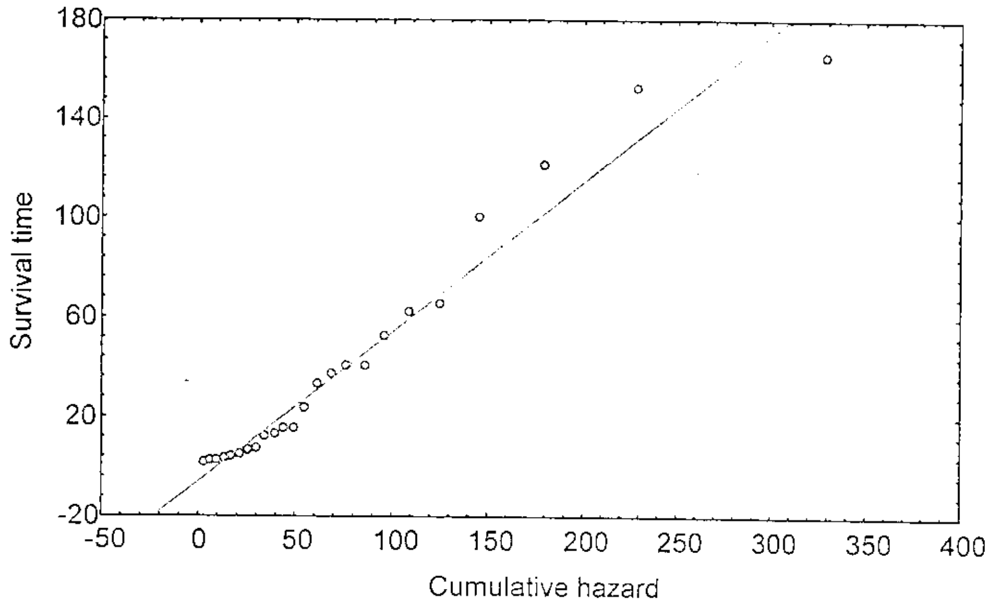
أولا : للمرضى المصابين بسرطان الرئة

X1بالأسابيع	ti بالأسابيع	Reverse Rank	Hazard Value	Cumulative Hazard Value
23	1	30	3.334	3.334
33	2	29	3.448	6.782
40	2	28	3.571	10.353
13	3	27	3.703	14.056
6	4	26	3.846	17.902
2	5	25	4.000	21.902
1	6	24	4.167	26.069
5	7	23	4.348	30.417
65	12	22	4.545	34.962
152	13	21	4.762	39.724
+39	15	20	5.000	44.724
4	15	19	5.263	49.987
+109	23	18	5.556	55.543
121	+30	17	—	—
15	33	16	6.25	61.793
12	+36	15	—	—
100	37	14	7.143	68.936
165	+39	13	—	—
+62	40	12	8.333	77.269
7	40	11	9.091	86.36
+55	52	10	10.000	96.36
62	+55	9	—	—
+30	62	8	12.500	108.86
2	+62	7	—	—
52	65	6	16.667	125.527
40	100	5	20.000	145.527
15	+109	4	—	—
+36	121	3	33.333	178.86
3	152	2	50.000	228.86
37	165	1	100.000	328.86

ثانياً: للمرضى المصابين بسرطان المعدة

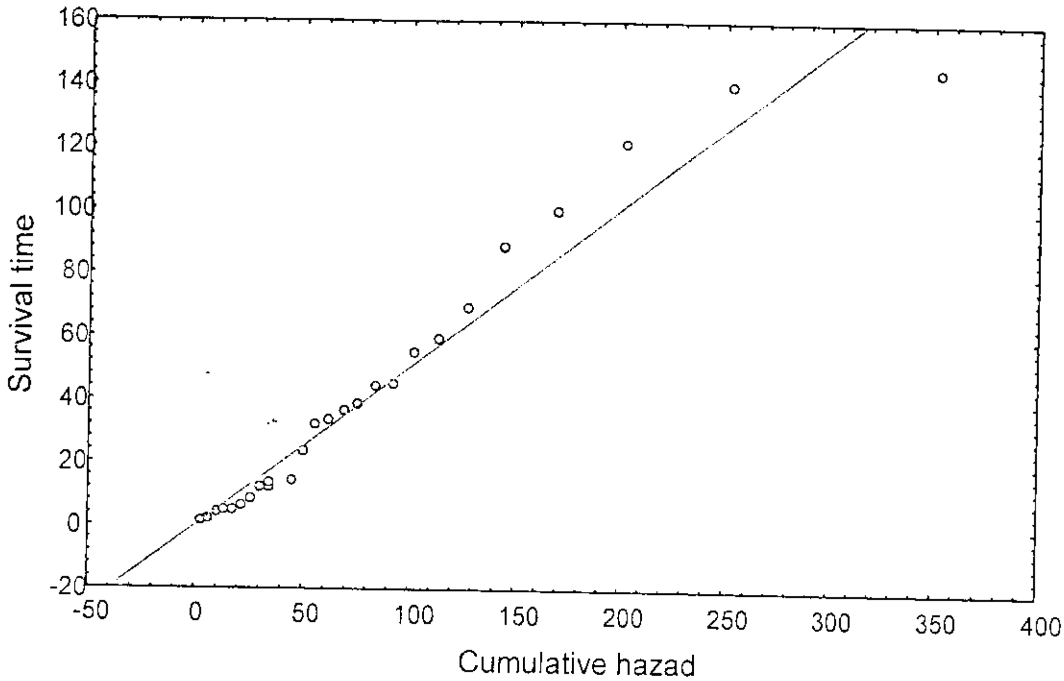
x2 بالأسابيع	ti بالأسابيع	Reverse Rank	Hazard Value	Cumulative Hazard Value
+55	1	30	3.334	3.334
145	2	29	3.448	6.782
+98	4	28	3.571	10.353
12	5	27	3.703	14.056
+14	5	26	3.846	17.902
100	6	25	4.000	21.902
121	8	24	4.167	26.069
4	12	23	4.348	30.417
38	12	22	4.545	34.962
140	13	21	4.762	34.724
36	+14	20	——	——
6	14	19	5.263	44.987
13	23	18	5.556	50.543
1	32	17	5.882	56.425
2	33	16	6.250	62.675
5	36	15	6.667	69.342
+65	38	14	7.143	76.485
44	44	13	7.692	84.177
32	45	12	8.333	92.51
23	+55	11	——	——
64	59	10	10.000	102.51
69	64	9	11.111	113.621
88	+65	8	——	——
5	69	7	14.286	127.907
45	88	6	16.667	144.574
59	+98	5	——	——
12	100	4	25.000	169.574
33	121	3	33.333	202.907
8	140	2	50.000	252.907
14	145	1	.100.000	352.907

علامة + تعني أن المفردة بقيت ما بعد الفترة ti



شكل (1)

دالة الخطورة التجميعية لمرضى سرطان الرئة



شكل (2)

دالة الخطورة التجميعية لمرضى سرطان المعدة

وبتطبيق المعادلة رقم 8 نجد إن $\hat{\lambda}_1 = 0.0167$ و $\hat{\lambda}_2 = 0.0175$ و بتطبيق المعادلة رقم 9 نجد إن $\hat{\lambda} = 0.0171$

ولاحتساب الإحصاءة Λ من المعادلة رقم 1 نجد إن $\Lambda = 0.967$ أما قيمة $-2Log_e\Lambda = 0.0671$ وبمقارنتها مع قيمة χ^2 الجد ولية بدرجة حرية 1بمستوى معنوية 0.01 والمساوية إلى 6.634 نجد إن القيمة المحسوبة اقل بكثير من القيمة الجد ولية وبهذا لا يمكن رفض فرضية العدم .

وباستخدام اختبار Cox F نجد أن $\bar{t}_1/\bar{t}_2 = 1.047$ وعند مقارنة القيمة المحسوبة مع قيمة F الجد ولية بمستوى معنوية 0.01 ودرجة حرية 48 للبسط و52 للمقام المساوية إلى 1.95 نجد أن القيمة المحسوبة اقل من الجد ولية ولا يمكن رفض فرضية العدم.

أهم استنتاجات البحث

لقد كشف اختبار Nelson إن بيانات كلا العينتين تتبعان التوزيع الاسي , وعلى الرغم من استخدام اختبار نسبة الإمكان الأعظم واختبار Cox F إلا أن كلا الاختبارين اثبتا إن الفروق غير معنوية للمرضى المصابين بسرطان الرئة والمرضى المصابين بسرطان المعدة وقد يعود السبب في ذلك أن كلا العينتين أُخذت للحالات المتأخرة من المرض أو انه فعلا كلا المرضيين هما بنفس الخطورة وهذا لا يمكن التأكد منة إلا إذ توفرت معلومات دقيقة في سجلات المرضى اذ تفتقر هذه السجلات الى الدقة المطلوبة لمثل هكذا الدراسات .

المصادر:

1- Cox, D.R.(1953), Some Simple Test for Poisson Variates,Biometrika, Vol. 40,354-360 .

2-Nelson,W.(1972), Theory and Application of Hazard Plotting for Censored Failure Data ,Technometrics,Vol.14,pp. 945-965.

3-Paul, W.Dickman & Timo Hakulinen.(2000), Survival Analysis, Doctoral Programs in Public Health, University of Helsinki.

....

....

....